

ON THE EXISTENCE OF LEAST FAVORABLE DISTRIBUTIONS¹

BY E. L. LEHMANN

Stanford University and University of California, Berkeley

1. Summary. Sufficient conditions for the existence of a least favorable distribution were given by Wald in his work on general decision theory. It is shown here that for problems of hypothesis testing and more generally for multiple decision problems involving a finite number of decisions, the result holds under a much weaker restriction than Wald's assumption of a compact parameter space.

2. Introduction. In his general theory of decision functions Wald has given conditions under which there exists a least favorable distribution for Nature. One of these, which is rather restrictive and not required for many parts of the theory, is condition (3.7) of [1], a compactness assumption on the parameter space. This condition can, of course, in general not be dispensed with; there are many examples of simple statistical decision problems for which a least favorable distribution does not exist. However, as we shall show below, for certain special classes of problems the compactness assumption is not required.

The following is a typical example. Suppose that θ is a location parameter and that we are interested in testing $H: \theta \leq \theta_0$ against $K: \theta \geq \theta_1$. If a least favorable distribution does not exist this means essentially that positive probability is assigned to at least one of the points $\theta = \pm \infty$. But it is intuitively clear that by doing this, Nature would be playing into the hands of the statistician, since for sufficiently large $|\theta|$ it is very easy to determine whether or not H is true. In fact one would expect Nature to do better if the above strategy S were replaced by the (conditional) distribution for θ obtained from S , given that no probability is placed on $\theta = \pm \infty$.

In the next section we shall make this argument precise and use it to establish the existence of a least favorable distribution for certain hypothesis testing and multiple decision problems. In particular our condition is satisfied for many problems of testing a composite hypothesis against a simple alternative. This shows that a method given in [2] for determining the most powerful test against a simple alternative is applicable in most problems of the kind usually considered.

3. Least favorable distributions for problems of hypothesis testing. In the present section we shall consider the hypothesis H that the probability density of a random variable X is $f(x)$ with $f \in \mathfrak{F}$, where \mathfrak{F} is some given class of densities, and the class of alternatives K that the density of X is $g_\theta(x)$ with $\theta \in \Omega$. We shall assume that the sample space \mathfrak{X} is Euclidean and that $\theta = (\theta_1, \dots, \theta_s)$, where Ω is a Borel set in R_s . Since throughout we shall be concerned with dis-

¹ Work sponsored in part by the office of Naval Research.

tributions over Ω we may, by defining these distributions to be zero outside of Ω , assume without loss of generality that Ω is the full Euclidean space R_n . We assume further that for almost all x , $g_\theta(x)$ is continuous in θ and measurable in $\Omega \times \mathfrak{X}$.

For the present we consider the usual asymmetric formulation of the problem in which a level of significance α is assigned and it is desired to maximize the minimum power of the test over K . In this situation, as is pointed out on page 133 of [1], assumptions (3.1) – (3.6) of [1] hold, and we have, as a simple consequence of Theorem 3.12 of [1],

LEMMA 3.1. *There exists a level- α test ϕ of H that maximizes the minimum power against K , and a sequence of distributions λ_i over Ω such that*

$$(3.1) \quad \lim_{i \rightarrow \infty} \int E_\theta \phi(X) d\lambda_i(\theta) = \inf_{\theta \in \Omega} E_\theta \phi(X)$$

and

$$(3.2) \quad \lim_{i \rightarrow \infty} \beta_i = \inf_{\theta \in \Omega} E_\theta \phi(X),$$

where β_i is the power of the most powerful level- α test for testing H against the simple alternative

$$\int g_\theta(x) d\lambda_i(\theta).$$

Now by Helly's compactness theorem there exists a subsequence of the sequence $\{\lambda_i\}$, which without loss of generality we shall take to be the original sequence, such that

$$(3.3) \quad \lambda_i \rightarrow \gamma \mu$$

for all continuity intervals of μ , where μ is again a distribution and $0 \leq \gamma \leq 1$. If here we could assume that $\gamma = 1$ it would follow that μ is least favorable (see the proof of Theorem 3.14 of [1]). We shall now show that μ is least favorable by proving that either $\gamma = 1$ or $\lim \beta_i = 1$ under the following

ASSUMPTION A. *Given any $\epsilon > 0$ and any closed bounded set $\omega \subset \Omega$ there exists a Borel set S in the sample space such that*

$$(3.4) \quad \begin{aligned} P(\bar{S} | f) &\leq \epsilon && \text{for all } f \in \mathfrak{F}, \\ P(\bar{S} | g_\theta) &\leq \epsilon && \text{for all } \theta \in \omega, \end{aligned}$$

and

$$(3.5) \quad P(S | g_\theta) \rightarrow 0 \text{ uniformly as } \theta_1^2 + \dots + \theta_n^2 \rightarrow \infty.$$

Clearly Assumption A is satisfied if both \mathfrak{F} and Ω are compact, but as we shall show later the assumption also holds in many cases in which (3.7) of [1] is not satisfied.

The proof of our result is based on the following two lemmas.

LEMMA 3.2. *Let ω be any closed bounded set in Ω and let $\epsilon > 0$. Then there exists a Borel set S in \mathfrak{X} and an integer i_0 such that*

$$(3.6) \quad \int_{\bar{\omega}} P(\bar{S} \mid \theta) d\lambda_i(\theta) \geq 1 - \gamma - \epsilon \quad \text{for } i \geq i_0,$$

and even

$$(3.7) \quad \int_{\bar{\omega}} E_{\theta}[\phi(X)c_S(X)] d\lambda_i(\theta) \geq 1 - \gamma - 2\epsilon \quad \text{for } i \geq i_0,$$

where ϕ is the test referred to in Lemma 3.1 and where $c_S(x)$ is the characteristic set function of \bar{S} .

Here and in what follows the notation $P(\bar{S} \mid \theta)$ and E_{θ} indicates that the probability or expectation is computed with respect to the probability density $g_{\theta}(x)$.

LEMMA 3.3. *For any $\epsilon > 0$ there exists a continuity interval ω of μ such that for any test ϕ*

$$(3.8) \quad \int_{\omega} E_{\theta} \phi(X) d\lambda_i(\theta) \geq \gamma \int_{\Omega} E_{\theta} \phi(X) d\mu(\theta) - \epsilon$$

for sufficiently large i .

Before proving these lemmas we shall show that they imply $\gamma = 1$ or $\lim \beta_i = 1$ and hence the existence of a least favorable distribution. Let ϕ be the test referred to in Lemma 3.1 and let

$$(3.9) \quad \lim_{i \rightarrow \infty} \int E_{\theta} \phi(X) d\lambda_i(\theta) = \beta.$$

Given any $\epsilon > 0$ let ω be a bounded subset of Ω for which (3.8) holds. Then

$$\int_{\Omega} E_{\theta} \phi(X) d\lambda_i(\theta) = \int_{\omega} E_{\theta} \phi(X) d\lambda_i(\theta) + \int_{\bar{\omega}} E_{\theta} \phi(X) d\lambda_i(\theta),$$

where for sufficiently large i the first term is $\geq \gamma \int_{\Omega} E_{\theta} \phi(X) d\mu(\theta) - \epsilon$ by (3.8) and the second term is $\geq 1 - \gamma - 2\epsilon$ by (3.7), so that

$$(3.10) \quad \beta \geq (1 - \gamma) + \gamma \int_{\Omega} E_{\theta} \phi(X) d\mu(\theta) \geq \int_{\Omega} E_{\theta} \phi(X) d\mu(\theta).$$

Here the last inequality is strict and therefore contradicts (3.1), unless either

$$\gamma = 1 \text{ or } \int_{\Omega} E_{\theta} \phi(X) d\mu(\theta) = 1.$$

But in the latter case $\beta = 1$ by (3.10), so that

$$\int_{\Omega} E_{\theta} \phi(X) d\mu(\theta) = \inf E_{\theta} \phi(X)$$

and μ would be least favorable.

PROOF OF LEMMA 3.2. Let S be a set guaranteed by Assumption A for which

$$(3.11) \quad P(\bar{S} \mid \theta) \leq \epsilon/2 \quad \text{for all } \theta \in \omega,$$

$$(3.12) \quad P(\bar{S} \mid f) \leq \epsilon \quad \text{for all } f \in \mathfrak{F},$$

and such that

$$(3.13) \quad P(S \mid \theta) \rightarrow 0 \text{ uniformly as } \theta_1^2 + \cdots + \theta_s^2 \rightarrow \infty.$$

From (3.3) and (3.13) it follows that

$$(3.14) \quad \int_{\Omega} P(\bar{S} \mid \theta) d\lambda_i(\theta) \geq 1 - \gamma - \epsilon/2 \quad \text{for } i \geq i_0,$$

and (3.6) is an immediate consequence of (3.14) and (3.11).

Suppose now that (3.7) did not hold. Then there would exist a subsequence of $\{\lambda_i\}$, which we may again take to be the full sequence such that

$$(3.15) \quad \lim_{i \rightarrow \infty} \int_{\bar{\omega}} E_{\theta}[\phi(X)c_S(X)] d\lambda_i(\theta) = \delta < 1 - \gamma.$$

We define a new test ϕ^* by

$$(3.16) \quad \phi^*(x) = \begin{cases} 1 & \text{if } x \in \bar{S}, \\ \frac{\alpha - \epsilon}{\alpha} \phi(x) & \text{if } x \in S. \end{cases}$$

Then for any $f \in \mathfrak{F}$ we have, because of (3.12)

$$(3.17) \quad E(\phi^*(X) \mid f) \leq \frac{\alpha - \epsilon}{\alpha} \alpha + P(\bar{S} \mid f) \leq \alpha,$$

so that ϕ^* is again a level- α test. Also, for all x ,

$$(3.18) \quad \phi^*(x) \geq \frac{\alpha - \epsilon}{\alpha} \phi(x) + c_S(x)[1 - \phi(x)],$$

and hence

$$(3.19) \quad \begin{aligned} \int E_{\theta} \phi^*(X) d\lambda_i(\theta) &\geq \frac{\alpha - \epsilon}{\alpha} \int E_{\theta} \phi(X) d\lambda_i(\theta) \\ &\quad + \int E_{\theta} \{c_S(X)[1 - \phi(X)]\} d\lambda_i(\theta). \end{aligned}$$

The first term on the right-hand side of (3.19) tends to $\beta(\alpha - \epsilon)/\alpha$ as $i \rightarrow \infty$. On the other hand the second term is greater than or equal to

$$\int_{\bar{\omega}} P_{\theta}(\bar{S}) d\lambda_i(\theta) - \int_{\bar{\omega}} E_{\theta}[c_S(X)\phi(X)] d\lambda_i(\theta),$$

which for sufficiently large i exceeds $(1 - \gamma - \epsilon) - (\delta + \epsilon)$ by (3.15) and (3.6). This shows that for sufficiently large i

$$\int E_{\theta} \phi^*(X) d\lambda_i(\theta) > \beta + \frac{1 - \gamma - \delta}{2}$$

and thus contradicts (3.2), since it implies that for sufficiently large i ϕ^* is more powerful than the most powerful test for testing H against $\int g_{\theta}(x) d\lambda_i(\theta)$.

PROOF OF LEMMA 3.3. Let ω be a continuity interval of μ , and so large that

$$(3.20) \quad \mu(\bar{\omega}) \leq \epsilon/2.$$

Then by the Helly-Bray theorem, as $i \rightarrow \infty$

$$\int_{\omega} g_{\theta}(x) d\lambda_i(\theta) \rightarrow \gamma \int_{\omega} g_{\theta}(x) d\mu(\theta),$$

and hence by Fatou's lemma

$$\liminf_i \int \left[\phi(x) \int_{\omega} g_{\theta}(x) d\lambda_i(\theta) \right] dx \geq \gamma \int \left[\phi(x) \int_{\omega} g_{\theta}(x) d\mu(\theta) \right] dx.$$

Interchanging the order of integration we then have for i sufficiently large

$$\int_{\omega} E_{\theta} \phi(X) d\lambda_i(\theta) \geq \gamma \int_{\omega} E_{\theta} \phi(X) d\mu(\theta) - \epsilon/2 \geq \gamma \int_{\Omega} E_{\theta} \phi(X) d\mu(\theta) - \epsilon$$

by (3.20).

We can sum up the work of this section in

THEOREM 3.1. *If $g_{\theta}(x)$ is continuous in θ and if Assumption A holds, then for the problem of testing H against K there exists a least favorable distribution for θ .*

As an illustration, let X_1, \dots, X_n be a sample from a distribution with density $f(x - \theta)$ and consider testing $H: \theta \leq \theta_0$ against $\theta \geq \theta_1$. Here we could clearly take for S a set of the form $X_1 \leq c$, so that Assumption A is satisfied.

4. Testing against a simple alternative. As an application of Theorem 3.1 we now consider the problem of testing a composite hypothesis against a simple alternative. Let the hypothesis state that the density of X is $f_{\theta}(x)$, $\theta \in \Omega$, and let the alternative be $g(x)$, where we assume as before that the sample space is Euclidean and that $\theta = (\theta_1, \dots, \theta_s)$ with Ω being the full Euclidean space R_s .

ASSUMPTION B. *For every closed bounded set S in the sample space, $P(S | f_{\theta}) \rightarrow 0$ uniformly as $\theta_1^2 + \dots + \theta_s^2 \rightarrow \infty$.*

THEOREM 4.1. *If $f_{\theta}(x)$ is a continuous function of θ for almost all x , then under Assumption B there exists a least favorable distribution over Ω .*

PROOF. Theorem 3.1 clearly applies if we interchange H , α with K , β . We therefore only need to show that Assumption B implies Assumption A with f and g interchanged. That is, we need to prove that Assumption B implies the existence of a set S such that

$$(4.1) \quad P(\bar{S} | g) \leq \epsilon,$$

$$(4.2) \quad P(\bar{S} | f_{\theta}) \leq \epsilon \quad \text{for all } \theta \in \omega,$$

and

$$(4.3) \quad P(S | f_\theta) \rightarrow 0 \text{ as } \theta_1^2 + \cdots + \theta_s^2 \rightarrow \infty.$$

It was pointed out by Scheffe [3] that when $g_\theta(x)$ is continuous in θ , the convergence $\theta_i \rightarrow \theta$ implies

$$(4.4) \quad \int_S g_{\theta_i}(x) dx \rightarrow \int_S g_\theta(x) dx \text{ uniformly for all Borel sets } S.$$

Therefore if ω is bounded and closed, the associated set of distributions is compact with respect to the convergence definition (4.4). Let $\epsilon > 0$. Then we can find a finite number of points in ω , say $\theta_1, \dots, \theta_r$, such that every $\theta \in \omega$ is within $\frac{1}{2}\epsilon$ of one of these θ_i in the sense that $|P(S | \theta) - P(S | \theta_i)| \leq \frac{1}{2}\epsilon$ for all Borel sets S . It follows that there exists a closed bounded set S for which (4.1) and (4.2) hold and (4.3) then follows from Assumption B.

ASSUMPTION C. *There exist s statistics $Y_i = h_i(X)$, $i = 1, \dots, s$, such that the h_i are continuous, the marginal distribution of Y_i depends only on θ_i , and Y_i tends to $+\infty$ or $-\infty$ as $\theta_i \rightarrow +\infty$ or $-\infty$.*

LEMMA 4.1. *A sufficient condition for Assumption B to hold is Assumption C.*

PROOF. Let S be a closed bounded set, and let $a_i = \min_S h_i(x)$, $b_i = \max_S h_i(x)$. Then $P_\theta(X \in S) \leq P_{\theta_i}(a_i \leq Y_i \leq b_i)$ for $i = 1, \dots, s$. Now if $\theta_1^2 + \cdots + \theta_s^2 > C$ we must have $\theta_i^2 > C/s$ for at least one value of i . Given $\epsilon > 0$, let C be so large that

$$P_{\theta_i}(a_i \leq Y_i \leq b_i) < \epsilon \quad \text{if} \quad \theta_i^2 > \frac{C}{s} \quad \text{for all } i.$$

Then $P_\theta(X \in S) \leq \epsilon$ provided $\theta_1^2 + \cdots + \theta_s^2 > C$.

As an example let X_1, \dots, X_n be a sample from a distribution with unknown location and scale parameter ξ and η . Let $\theta_1 = \xi$, $\theta_2 = \log \eta$, and take $Y_1 = X_1/|X_2 - X_1|$, $Y_2 = \log |X_2 - X_1|$. Then Assumption C is seen to be satisfied, so that for testing a hypothesis of this type against a simple alternative, a least favorable distribution exists.

Theorem 4.1 answers a problem raised by [2], where in Theorem 1 a method is given for proving that a test is most powerful for testing a composite hypothesis against a simple alternative. Combining Theorem 4.1 with Theorem 3.10 of [1], we see that this method is always applicable when Assumption B is satisfied.

While it is quite likely that Assumption B can be weakened somewhat, it is perhaps of interest to point out that some condition of this type is required, which restricts the behavior of the distributions as $\theta_1^2 + \cdots + \theta_s^2 \rightarrow \infty$. As an example consider the following situation in which Ω is denumerable and the densities are given by

$$H: f_i(x) = \frac{1}{2\sqrt{2\pi}} [e^{-\frac{1}{2}(x-i)^2} + e^{-\frac{1}{2}(x-(1/i))^2}], \quad i = m, m+1, \dots,$$

$$K: g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

We shall show that if m is sufficiently large the most powerful level- α test rejects when $x \leq b$, where b is determined by

$$(4.5) \quad \lim_{i \rightarrow \infty} \int_{-\infty}^b f_i(x) dx = \frac{1}{2} \int_{-\infty}^b g(x) dx = \alpha.$$

First one checks easily that for large i $\alpha(i) = \int_{-\infty}^b f_i(x) dx$ is an increasing function of i . Let m be such that $\alpha(i)$ is increasing for $i \geq m$. Then the proposed test is of size α and power 2α by (4.5). But if β is the power of the most powerful test ϕ we clearly have

$$\beta \leq 2 \limsup_{i \rightarrow \infty} \int f_i(x) \phi(x) dx \leq 2\alpha,$$

which proves that the region of rejection $x \leq b$ is most powerful. If a least favorable distribution λ over Ω existed, the test $x \leq b$ would be a Bayes solution for λ (Theorem 3.9 of [1]), that is, would be most powerful for testing $\int f_i(x) d\lambda(i)$ against $g(x)$. But an application of the Neyman-Pearson fundamental lemma shows immediately that no Bayes solution is of this form.

It may be worth pointing out that the conditions of the present paper are really of a quite different nature from those given by Wald in his general theory. Thus they do not even imply that the parameter space is weakly compact in the sense of (3.1) of [1]. As an example, consider a random variable X the distribution of which depends on an unknown location parameter θ . Let H denote the hypothesis $\theta \leq \theta_0$ and suppose that the simple alternative is $\theta_1 > \theta_0$. Let $\{\lambda_i\}$ be any sequence of distributions over H , and let $\phi(x)$ be any test. Then weak compactness of H would imply the existence of a subsequence $\{\lambda_{i_j}\}$ and a distribution λ_0 over H such that

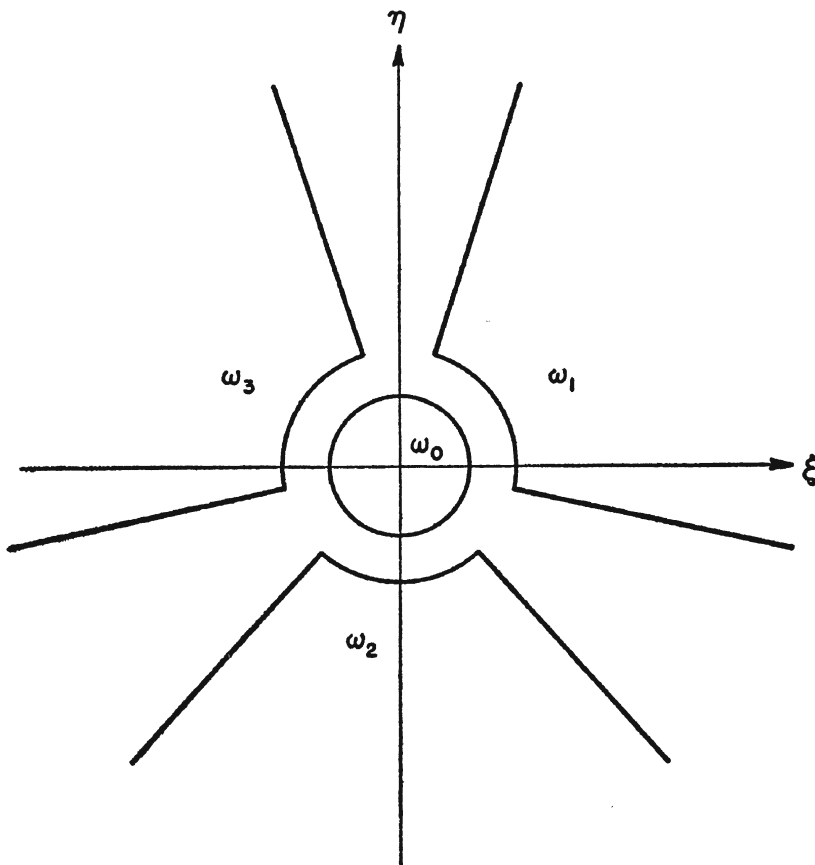
$$(4.6) \quad \limsup_{j \rightarrow \infty} \int E_{\theta} \phi(x) d\lambda_{i_j}(\theta) \leq \int E_{\theta} \phi(x) d\lambda_0(\theta).$$

But let $\phi(x)$ be the characteristic set function of the set $x \leq C$ and let $\lambda_i(\theta)$ assign probability 1 to the point $\theta = -i$. Then the left-hand side of (4.6) is 1, while for any distribution λ_0 the right-hand side is < 1 , so that H is not weakly compact. On the other hand, Assumption B is satisfied.

5. Some extensions. We shall now indicate briefly how the results of Section 3 may be extended to somewhat more general decision problems. As a first generalization consider the problem of hypothesis testing, formulated as in Section 3 except that the loss function need no longer be simple. Suppose that acceptance of the hypothesis when the true distribution is given by $g_{\theta}(x)$, $\theta \in \Omega$ results in a loss $W(\theta)$, and that $V(f)$ is the loss resulting from rejection of H when f is the true density. We shall assume that V and W are bounded and that we are interested in minimizing $\sup_{\theta} W(\theta) E_{\theta}[1 - \phi(x)]$ among all tests ϕ for which $V(f) E[\phi(x) | f] \leq \alpha$ for all $f \in \mathfrak{F}$.

Let Assumption A' be obtained from Assumption A by replacing $P(\bar{S} | g_\theta)$ and $P(S | g_\theta)$ in (3.4) and (3.5) by $W(\theta)P(\bar{S} | g_\theta)$ and $W(\theta)P(S | g_\theta)$ respectively, and $P(\bar{S} | f)$ in (3.4) by $V(f)P(\bar{S} | f)$. Then Assumption A' implies the existence of a least favorable distribution over Ω . Only the obvious changes are required in the proof of Theorem 3.1.

As an immediate further generalization one next obtains an analogous result for the following two-decision problem. The parameter space is partitioned into two set $\omega_1 \cup \omega_2$. The loss is zero if $\theta \in \omega_i$ and decision d_i is taken; it is $W_i(\theta)$ if



decision d_i is taken incorrectly. Let us denote the risk function of a decision procedure δ as usual by $R_\delta(\theta)$ and suppose that if δ is the minimax solution

$$\sup_{\omega_i} R_\delta(\theta) = \gamma.$$

Then the overall minimax problem is equivalent (see [4]) to the problem Π_1 of minimizing $\sup_{\theta \in \omega_1} R_\delta(\theta)$ subject to $R_\delta(\theta) \leq \gamma$ for all $\theta \in \omega_2$, and also to the problem Π_2 obtained from Π_1 by interchanging 1 and 2. If a least favorable distribution over ω_i exists for Π_i ($i = 1, 2$), then one exists for the overall problem. But each of the partial problems is of the type discussed at the beginning of this section.

The whole argument extends without much difficulty to multiple decision

problems involving a finite number of decisions and having a bounded loss function. We shall not give the details but simply state the result. Let $\omega_1, \dots, \omega_K$ be the regions in the parameter space in which the decisions d_1, \dots, d_K are appropriate so that the loss $W(\theta, d_i) = 0$ if $\theta \in \omega_i$. Then we consider the two-decision problems Π_i ($i = 1, \dots, K$), where the decision lies only between $\theta \in \omega_i$ and $\theta \in \Omega - \omega_i$. A least favorable distribution exists over Ω provided each of the problems Π_i satisfies Assumption A'.

As an example consider the case that $X_1, \dots, X_m; Y_1, \dots, Y_n$ are $m + n$ independent random variables, and that the density of the X 's is $f(x - \xi)$ while that of the Y 's is $g(y - \eta)$. Suppose that the partition of Ω is as shown in the diagram and that the loss is 0 or 1 as the correct or an incorrect decision is taken. Then a least favorable distribution exists. To show that Π_1 satisfies Assumption A' suppose that the bounding rays l_1 and l_2 of ω_1 are given by $\eta = m_1\xi$ and $\eta = m_2\xi$. Then for \tilde{S} we may take a set in the sample space of the form

$$x_1^2 + y_1^2 \geq R, \quad m_1 + c_1 \leq y_1 \leq m_2x_1 + c_2,$$

where R , c_1 and c_2 must be chosen sufficiently large.

REFERENCES

- [1] A. WALD, *Statistical Decision Functions*, John Wiley & Sons, 1950.
- [2] E. L. LEHMANN AND C. STEIN, "Most powerful tests of composite hypotheses. I. Normal distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 495-516.
- [3] H. SCHEFFÉ, "A useful convergence theorem for probability distributions," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 434-438.
- [4] C. BLYTH, "On minimax statistical decision procedures and their admissibility," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 22-42.